

A Data-Driven Approach for Semantic Role Labeling from Induced Grammar Structures in Language

Vivek Datla¹, David Lin², Max Louwerse³,
Abhinav Vishnu⁴

¹Philips Research, Cambridge, MA; ²Baylor University, Waco, TX;

³Tilburg University, Tilburg, Netherlands; ⁴Pacific Northwest National Laboratory, Richland, WA
vivek.datla@philips.com, davidlin@baylor.edu,
maxlouwerse@tilburg.edu, abhinav.vishnu@pnnl.gov

Abstract

Semantic roles play an important role in extracting knowledge from text. Current unsupervised approaches utilize features from grammar structures, to induce semantic roles. The dependence on these grammars, however, makes it difficult to adapt to noisy and new languages. In this paper we develop a data-driven approach to identifying semantic roles, the approach is entirely unsupervised up to the point where rules need to be learned to identify the position the semantic role occurs. Specifically we develop a modified-ADIOS algorithm based on ADIOS Solan et al. (2005) to learn grammar structures, and use these grammar structures to learn the rules for identifying the semantic roles based on the context in which the grammar structures appeared. The results obtained are comparable with the current state-of-art models that are inherently dependent on human annotated data.

Introduction

For speakers and hearers to understand the meaning of the message, they need to understand the “who-did-what-to-whom”, the semantic roles within a sentence. These semantic roles are used to denote the underlying relationship between various constituents of a sentence, mostly with the main predicate (verb) of the sentence. For example, in the sentence “*John kissed Mary*”, we denote “*John*” as the “Agent” and “*Mary*” as the “Patient” with respect to the relation to main verb “*kissed*”.

Semantic roles play a crucial role in a variety of NLP applications because they provide crucial information to the meaning of the individual words. The role of “*Mary*” is different when *Mary* is a Patient (“*John kissed Mary*”) or an Agent (“*Mary kissed John*”). Note that the syntactic Subject or Object might be different while the Agent and Patient may remain the same (e.g., “*John kissed Mary*” and “*Mary was kissed by John*”). NLP applications in which semantic roles are frequently used include information extraction, question answering, automatic summarization, and many more.

One can view SRL as a two-step process: 1) Identifying the slots in a sentence that are most likely to contain the semantic roles, and 2) Selecting a specific semantic role that mostly likely fits those slots. Computationally both these problems are framed as machine learning models.

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Table 1: Human annotated data dependency among several components of unsupervised SRL systems; part-of-speech(1); parsers(2); slot for semantic role(3); labeling semantic role(4); Yes(Y), No(N), partial(Y_p)

Literature	1	2	3	4
Abend, Reichart, and Rappoport (2009)	Y	Y	Y	N
Garg and Henderson (2012)	Y	Y	Y	N
Grenager and Manning (2006)	Y	Y	Y_p	Y_p
Lang and Lapata (2010)	Y	Y	Y	N
Lang and Lapata (2011)	Y	Y	Y	N
Modi, Titov, and Klementiev (2012)	Y	Y	Y	N
Titov and Klementiev (2011)	Y	Y	Y	N
our method	N	N	Y_p	Y_p

While there has been a lot of work on supervised SRL methods, supervision does have its limitations:

- Annotating gold standard data is a slow and resource intensive process
- Dealing with novel words or novel word combinations can create problems (“I googled about pizza today”); current models cannot handle mixed languages, making it easy for existing models to become obsolete item SRL models trained in one domain usually do not perform well in other domains Pradhan et al. (2005).

Exploring the effectiveness of using unsupervised learning methods for SRL is, therefore, a worthwhile exercise.

Several unsupervised approaches have been developed to extract semantic roles from text (see Table 1). These studies apply methodologies to exploit the structure of the language to extract semantic roles. The structures are built using syntactic parsers built from large human annotated corpora. Similar semantic roles tend to appear in the same part of the dependency tree, and unsupervised semantic role labeling models exploit this phenomenon. However, even though these methods do not need a large corpus for role labeling, they do require (large) human annotated data for building the structures (e.g. dependency tree, part-of-speech and constituent labels) for the labeling task. The current methods for unsupervised semantic role labeling still suffer from the limitations above. Table 1 indicates the supervised, unsupervised and semi-supervised components in various parts of semantic role labeling.

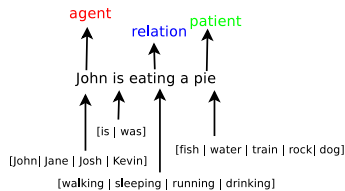


Figure 1: Thematic role of equivalent sentences

The main goal of the current paper is to build a model that extracts semantic roles present in language using bottom-up or data-driven approaches, without the need of human annotated data for structure building. Our method starts out by learning patterns of the language using re-occurring phrases within a context. We learn more complex patterns by increasing the context on existing patterns, and we cluster the patterns into groups based on word overlap among patterns. Specialization increases the content inside patterns and generalization helps to find equivalent patterns in a given context. Since patterns are learned from language, and patterns are organized hierarchically, identifying the semantic role in one of the patterns would help us percolate the semantic information to all parents of the pattern. This organization of patterns gives us the additional advantage of needing only a relatively small training set of known semantic roles (e.g. proportional to the number of patterns) for an effective SRL model.

Figure 1 illustrates our method. Starting with a sentence “John is eating a pie”, we build a pattern of “X is Y a Z” where X, Y, Z are slots that contain multiple terms that appear in sentences in our unannotated corpus that share the pattern. Here we can see the location of the agent and patient on the relation in that structure does not change. This consistency in the structure allows us to learn the semantic roles based on the patterns.

We modeled our pattern learning method based on a modified version of ADIOS algorithm Solan et al. (2005), one of the few algorithms that model the learning of language as bottom-up machine learning process. Following the description of the process how to create the patterns and how to use them for SRL process, we will present a computational study showing the success of our methods that do not require a large annotated corpus compared to alternative methods.

Data-driving SRL procedure

Our proposed semantic role labeling model consists of two steps:

- **Extracting patterns and rules from the text:** The first step is to take a *unannotated* corpus of text and discover common patterns and structures embedded in them. As mentioned, a pattern is not only a recurrence of statements but also adherence to a form. For example “X Y a Z” represents a form where the letters denote set of terms that can appear in that location. The terms may be a single word (e.g. X=John, Y=eats, Z=pie) or they can also be other patterns (e.g. Z = U made in V, where U=pie, V=France). Thus, the patterns are inherently organized to form hierarchical structures. Once we extract the patterns and struc-

tures, they are transformed into (production) rules. These rules are then used later for the parsing of the text in the SRL task.

- **Semantic role learning/labeling:** We use the rules learned to parse sentences of a given training set, generate features from the parsed sentences and use those features to train a classifier for identifying semantic roles. We use a small set of human annotated data to train the classifier.

The following sections present these steps in detail.

Extracting patterns and rules via Modified ADIOS(m-ADIOS)

To extract patterns and structures from the text, we apply a modified version of the Automatic Distillation of Structure (ADIOS) algorithm Solan et al. (2005). The algorithm builds generative rules based on the symbolic sequential information (i.e. language input as text).

In the ADIOS algorithm, a corpus is represented by a directed multigraph, where each vertex denotes a word. Each sentence is represented by a path connecting these words. If a phrase (e.g. John is) appears in multiple sentences, each occurrence will contribute a separate edge between the words (hence the multi-graph). Each sentence is complemented by a “begin” and an “end” vertex. The multigraph representation helps to identify those vertices that are thickly connected. They inherently represent the sequences that frequently appear in the corpus.

The algorithm iteratively picks a path (say v_1, \dots, v_n), and determines the part of the sentence that qualifies as a significant pattern. For each term (v_i) in the path, we calculate the conditional probability that a path from v_1 to v_{i-1} will continue to v_i (denoted as $P_R(v_1, v_i)$), and use a statistical test to detect the most significant drop in $P_R(v_1, v_i)$. This determines the right end of the pattern. A similar construct $P_L(v_n, v_i)$ – the conditional probability that sentences go from v_n to v_{i+1} would have come from v_i is used to find the left end of the pattern. Once the significant pattern is found, it is replaced by a single vertex and the graph is rewired.

Note that the new vertex does not replace the vertices in the path, but only consolidates all the paths represented by the pattern into a vertex – for example, if v_2, v_3, v_4 forms a significant pattern, then a vertex P is used to represent all paths that pass through the three vertices. v_2, v_3, v_4 are not removed, if there are other alternate routes between v_2, v_3, v_4 and other vertices. The single vertex that represents the pattern can also be looked at as a hierarchical structure. These hierarchical structures (new vertices) appear in a context. Consequently, we search through all the paths in which this vertices appear and select a significant path and generalize on the vertices that appear in the same context.

Furthermore, a generalization process of the patterns allows for consolidation of the patterns. If two patterns differ by only one vertex, we put all vertices into one class called the Equivalence class and replace all the vertices inside the equivalence class by a single vertex.

These steps are repeated until all the paths are iterated and no other significant patterns are formed. The output of

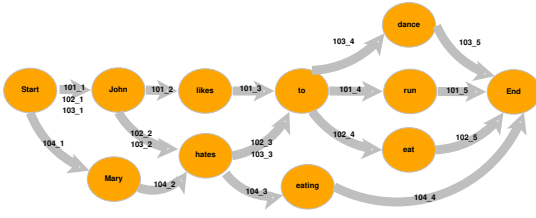


Figure 2: Data representation in modified-ADIOS

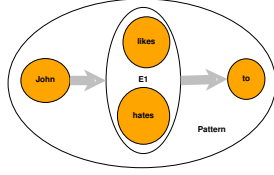


Figure 3: Pattern structure

this process consists of rules that are similar to context-free grammar(CFG) rules that will be used in subsequent steps.

Some modifications are made to the ADIOS algorithm. First we consolidate the representation of the graph by using a simple directed graph, but with edge labels denoting the various sentence corresponds to each edge. This modification in the structure makes the graph scalable. For example, Figure 2 shows the graph built on a corpus of four sentences, *John likes to run*, *John hates to eat*, *John hates to dance*, and *Mary hates eating*.

The ADIOS model generalizes greedily and allows for structures with variable sizes to be represented as one structure. This leads to phrases of various sizes not occurring in the same context to be considered as one structure, there by creating unwanted ambiguity in the grammar. Our goal here is not to produce a very compact grammar but a non-ambiguous grammar. We, therefore, modified the ADIOS algorithm to learn only the structures that occur within a context and have a very strict method for generalization.

We first create an equivalence class with all words that share the same left and right context. Then merge the equivalence class, and the respective left and right context to a single pattern. For example, in Figure 2 if “John likes to” is significant then in the context of “John” and “to” we find that “likes” and “hates” can be merged into one equivalence class. We merge both vertices and make an implicit note in our grammar tree that make the tokens “likes” and “hates” fall into the same equivalence class. Vertices “John”, “equivalence class” and “to” are then merged into a single vertex. The pattern learned from the text is shown in Figure 3.

It is important to note that we only replace these three vertices and two links and do not change any other vertices in the graph.

After finding a significant pattern, we check if we can find equivalent middle words for the same left and right context. By finding the middle words, which share the same context, we are more likely to find words which if interchanged also form syntactically correct sentences. After finding the middle words that share the same left and right context, we put them into a group called equivalence class.

Next, we check for the equivalence of a pattern. The

equivalence of a pattern is defined as the maximal overlap of the pattern with the existing patterns. Since each pattern consists of a left context, a right context, and an equivalence class, we consider two patterns to be equivalent if two of the three components match.

Every time we find a pattern we check if we can find an equivalence of that pattern to find structures that are replaceable. We repeat the process until we have covered all the paths and can no longer find any additional significant patterns. By following the above methodology, we end up with several equivalence classes that can be made up of patterns, words or other equivalence classes. A pattern can consist of words, equivalence classes and also other patterns.

We also generalize the equivalent patterns if they share at least left or right context and have more than σ overlap,

$$\text{i.e. } \frac{\text{size}(E_i \cap E_j)}{\text{size}(E_i \cup E_j)} \geq \sigma.$$

The algorithm 2 shows the steps followed for generalizing the equivalent classes. After finding the patterns, we consolidate all rules that we learned in the form of hierarchical structures. These hierarchical structures have the base patterns as leaf nodes, and patterns that use the base structures as parents of the base patterns and so on. As we move up the tree, we would be increasing the information in the pattern, as the size of the pattern keeps increasing as it subsumes the information from the nodes below. Even though the information increases, the nodes that are subsumed by this parent node would have the same dependency structure as the original base nodes. Since the structure is the same, identifying the semantic role slots in the base nodes would also be the slots for semantic roles in the parent nodes.

Algorithm 1 Build a sparse directed graph $G = (V, E)$ where each Sentence $s_i \in C$ is a path

Input: Sentences from Corpus C
Output: Sparse directed graph G

```

for each sentence  $s_i \in C$  do
  for each word  $w_n \in s_i$  do
    if  $w_n \notin G$  then
      add  $w_n$  to  $G$ 
    end if
    if  $w_{n+1} \notin G$  then
      add  $w_{n+1}$  to  $G$ 
    end if
    if An edge exists between  $w_n$  and  $w_{n+1}$  then
      Add sentenceID and linkID to the edge list.
    else
      Add directed edge from  $w_n \rightarrow w_{n+1}$  in Graph  $G$ 
      Add sentenceID and linkID to the edge list.
    end if
  end for
end for

```

At the end of m-ADIOS, the patterns and classes take the form of a hierarchical structure. For us to use them in subsequent steps, we would derive rules from them, so that we can then parse any statement through the rules for the semantic role labeling task. Each rule contains a node of the tree on

Algorithm 2 Generalization of Equivalence Patterns

```

 $E_i = P_i \rightarrow L_i E_i R_i$ 
 $E_j = P_j \rightarrow L_j E_j R_j$ 
if ( $L_i = L_j$ ) AND Equivalence class overlap ( $E_i, E_j$ )  $\geq \sigma$  then
    Build new Equivalence Class  $E_k = R_i \cup R_j$ 
    Merge Equivalence Patterns  $E_i$  and  $E_j \rightarrow E_{i,j}$ 
    Build new Pattern  $L_i, E_{i,j}, R_k \rightarrow P$ 
end if
if ( $R_i = R_j$ ) AND Equivalence class overlap ( $E_i, E_j$ )  $\geq \sigma$  then
    Build new Equivalence Class  $E_k = L_i \cup L_j$ 
    Merge Equivalence Patterns  $E_i$  and  $E_j \rightarrow E_{i,j}$ 
    Build new Pattern  $L_k, E_{i,j}, R_i \rightarrow P$ 
end if
if ( $L_i = L_j$ ) AND ( $R_i = R_j$ ) then
    Merge Equivalence Patterns  $E_i$  and  $E_j \rightarrow E_{i,j}$ 
    Build new Pattern  $L_i, E_{i,j}, R_i \rightarrow P$ 
end if

```

Table 2: CFG like rules generated using m-ADIOS

```

ROOT  $\rightarrow$  S
S  $\rightarrow$  john Pattern_118 mother
Pattern_118  $\rightarrow$  fedexed E_56 his
E_56  $\rightarrow$  Pattern_109
Pattern_109  $\rightarrow$  his E_55 to
E_55  $\rightarrow$  computer | classes | heart | horse | opposition |
socks | package | finger | cheeks | father |

```

the left and all its children in sequential fashion on the right. Table 2 shows an example of CFG like rules created using patterns, equivalent classes, and reduced paths in the graph.

Semantic role labeling

The semantic role labeling task involves learning from a corpus of sentences annotated with semantic roles. We treat each sentence as a single unit (as semantic roles relate terms from a sentence). We first parse each sentence, using the hierarchical structure rules learned from the m-ADIOS algorithm. After parsing the sentences, we locate patterns and collect contextual information around them. This information will be used as features for a classifier supporting learning and labeling purposes.

Data: Propbank We illustrate this process by using PropBank Kingsbury and Palmer (2002) as our corpus. PropBank annotated text from the Penn Treebank and the Wall Street Journal Corpus. PropBank is based on predicates, in the majority of the cases they are verbs. Each predicate has a set of arguments that is associated (from ARG0 to ARG5) associated to it. Arguments come with types such as location (LOC), temporal (TMP), manner (MNR), etc. The first two arguments, ARG0, and ARG1, are similar to prototypical agent and patient respectively. For the current purposes of the paper, we use only those sentences fewer than ten words. After filtering all 10600 sentences, we end up with 2249 sentences. The 2249 sentences from PropBank includes 4892

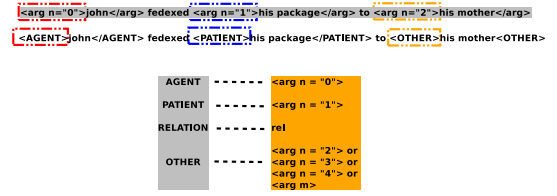


Figure 4: PropBank annotation scheme

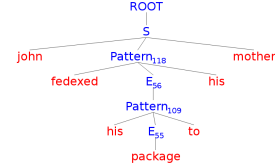


Figure 5: Parsing using grammar rules learnt from raw text using m-ADIOS

unique words which occur 20148 times (on average each word occurs 4.12 times (SD= 51.05)). In the current work, we concentrate only on the agent-patient-relation. Relation represents the verb for which agent and patient are semantic roles.

Figure 4 shows an example of an annotated sentence by PropBank, and also simplified annotation for our learning task. $\langle \text{arg } n = "0" \rangle$ represents an Agent present in the sentence, we identify this information using the tag "Agent", $\langle \text{arg } n = "1" \rangle$ represents the Patient, we identify this information using the tag "Patient", $\langle \text{arg } n = "2" \rangle$, $\langle \text{arg } n = "3" \rangle$, $\langle \text{arg } n = "4" \rangle$ and $\langle \text{arg } m \rangle$ represent meta tags, which add or modify information to Agent or Patient, we identify this information using the tag "Other" and $\langle \text{rel} \rangle$ indicates the relation or the action present in the sentence, we identify this information using the tag "Relation".

Feature extraction To annotate the data, we take each sentence of the Propbank annotated sentence, and then parse the sentences based on the rules generated in the previous step

For illustration purposes let's look at the sentence "John FedExed his package to his mother". This sentence when parsed gives us a tree shown in Figure 5. The syntactic tree using rules derived from language deviates from the parse done using Stanford parser Socher et al. (2013) (see Figure 6) which is considered a state-of-art parser. Even though it does not match the output of a standard parser, rules learned from language consistently parse similar sentences the same way. This has the advantage that - we can use the rules to train model for identifying semantic roles.

Once we parsed the sentence, we extract various features of the parse tree and use them as our features for the classification. The features we have used are shown in Table 3.

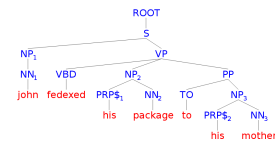


Figure 6: Parsing using Stanford Parser

Table 3: Features for classification

Features used for classification
1. Head of Phrase
2. Word two words before the pattern
3. Word one word before the pattern
4. Words inside the pattern
5. Word adjacent to the pattern
6. Word two words after the pattern
7. Length of the Pattern

Two words before and after the pattern are used to capture the context of the pattern. Pattern labeling is analogous to the head word, words inside the pattern are analogous to the path between the semantic roles, and length of the pattern gives information about the span of the pattern. The use of surface features of the parse tree to identify semantic roles is a standard practice in several supervised and unsupervised methodologies of semantic role labeling.

Labeling the data Since patterns and equivalent classes learned using the m-ADIOS algorithm does not have control over phrase boundaries, there are cases in which multiple semantic roles can be encompassed inside a pattern. For example, Figure 7 shows that “Pattern_118” encodes “Patient” and partially encodes “Other”. For simplifying the task, we assume that we encapsulate the semantic role even if we encapsulate a single word inside the semantic role. So we give the label “Patient_Other” for “Pattern_118”.

In this study we focused on identifying agent, patient and relation, we simplify our class labels by considering them as a triplet of boolean decisions. The triplet is made of “<Agent, Patient, Relation>”, so the class label for “Pattern_118” would be “<false,true,false>”. The classification task using the above triplet “<Agent, Patient, Relation>” is an 8 class classification with possibility of each entity being true or false.

Based on seven features shown in Table 3, and the class label given to each instance of the patterns, we do a classification via a 10-fold cross validation using Bayes, Naive Bayes, and Random Forest classifiers using Weka Hall et al. (2009).

Summary of Dataset Creation

To test the performance of m-ADIOS in the generating structure which encodes semantic roles, we use sentences in PropBank and parsed sentences obtained after parsing sentences using rules learned from the language. Let us look at the example closely, the sentence “john fedexed his package to his mother”, is parsed using the rules generated from the PropBank corpus. After parsing, we end up with a tree like structure, which is represented using a bracketing scheme.

Figure 5 shows tree representation of the same bracketing scheme. Figure 6 represents the structure using the Stanford parser. We can observe that patterns take the role of “head” similar to that of Noun Phrase (“NP”), Verb Phrase (“VP”) etc. Patterns learnt from the grammar rules do not have correct phrase boundaries in contrast to the strict adherence to phrase boundaries in traditional parsers (Figure 6).

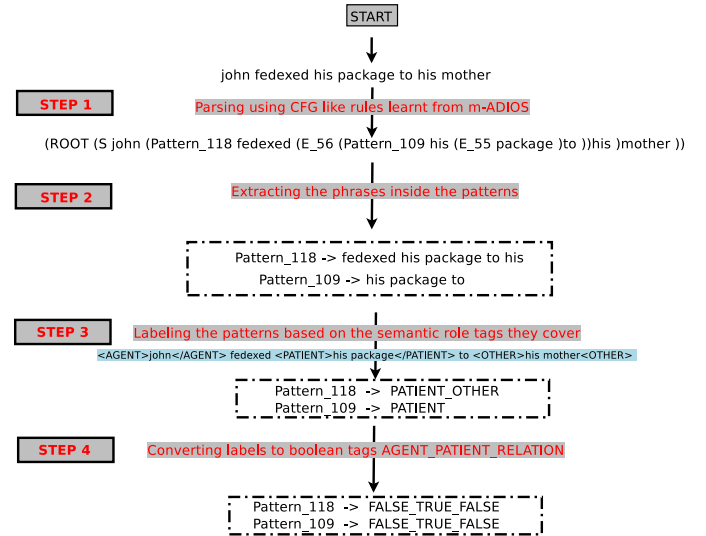


Figure 7: Creating dataset for classification

Since PropBank follows strict phrase boundaries, which adheres to Penn-tree bank annotations, our patterns transgress the defined phrase boundaries. As a result of these transgressions we some times have more than one role inside the pattern.

In step 3 in the Figure 7 we capture the semantic roles encapsulated by the patterns. Here we can see that “Pattern_118” encodes “Patient” and partially encodes “Other.” For simplifying the task, we assume that we encapsulate the semantic role even if we encapsulate a single word inside the semantic role. So we give the label “Patient_Other” for “Pattern_118”.

We perform these four steps on all the sentences for which we have learned rules using the m-ADIOS. The classification task using the above triplet “<Agent, Patient, Relation>” is an 8 class classification with a possibility of each entity being true or false. To summarize, given a pattern we would predict whether it encodes an agent or a patient or a relation or any combination of them.

Results and Discussion

Based on the 7 features shown in Table 3, and the class label given to each instance of the patterns following Section , we did a classification via a 10-fold cross validation using Bayes, Naive Bayes and Random Forest models. We report all the results using the measures precision $P = \frac{\text{No. of correct constituents identified}}{\text{Total no. of constituents identified}}$; recall $R = \frac{\text{No. of correct constituents identified}}{\text{Total no. of correct constituents}}$; f-measure $F = \frac{2 \times PR}{P+R}$ and inter-rater reliability using Cohen’s kappa K .

Table 4 shows the classification results. As expected the accuracy is reasonably high given that the chance of finding the correct tag is only 0.125. The classification is done with a great amount of confidence as well ($k > 0.6$). These can be attributed to the sparsity in the patterns obtained from the language.

Semantic roles are defined by order and kind of words present in the sentence, gave us the motivation to explore

Table 4: Results showing the classification for actual sentences

Classifier	P	R	F	K
BayesNet	0.81	0.795	0.789	0.711
NaiveBayes	0.807	0.767	0.749	0.664
Random Forest	0.79	0.75	0.75	0.64

Table 5: Results of state-of-art unsupervised semantic role labeling models

Work	Purity	Precision	F1
Lang and Lapata (2010)	0.80	0.77	0.78
Lang and Lapata (2011)	0.89	0.73	0.80

the distributional information around the words, and hierarchical relations among them. We further developed the m-ADIOS algorithm inspired from ADIOS Solan et al. (2005). In the m-ADIOS algorithm, the structural information learned is strictly dependent on the left and the right context in which words appear. We generalized this by assuming that only when two patterns have a perfect overlap on the right and left context or when there is a match with a left or right context of the two patterns and a significant overlap in their equivalent structures thereby reducing the ambiguity in the grammar.

Lang and Lapata (2010, 2011) used dependency trees from PropBank to extract features like predicate-lemma, argument-lemma, argument part-of-speech, and part-of-speech of left and rightmost child of argument to cluster the semantic roles in the sentences. Table 5 shows the baseline accuracies of current state-of-art unsupervised srl models. Even though their accuracy is high, their latent dependence of human annotated data dependency parsers, part-of-speech taggers makes it pseudo-unsupervised. Several other unsupervised models of semantic role labeling (see Table 1) also share the same issues.

Our results show that we can learn structure from language, in a data-driven fashion, and with little training we were able to identify the presence and absence of semantic roles. Even though the precision and recall scores may be lower when compared to existing unsupervised models, by not being dependent on human annotated data the proposed method would help to identify semantic roles for languages for which syntactic parsers developed from human annotated data are unavailable.

Future Work

The strict generalization in our model allows patterns that share a strong overlap in their context to be termed equivalent. After learning the rules, we convert them to grammar rules and parse the sentences from which we have learned the rules. In theory, these rules can be extended to any other sentences, which have the same words. But in practice this becomes a major drawback as we have the problem of sparsity simply because many words can occur in different contexts, and the order they appear in makes a big difference on the forming of patterns. To overcome this problem we plan to the method used by Datla, Lin, and Louwerse (2014) to induce pseudo parts of speech, and learn structures based on

the induces parts of speech.

In future we would like to extend our model for other languages, and also to the languages that are resource constraint.

References

- Abend, O.; Reichart, R.; and Rappoport, A. 2009. Unsupervised argument identification for semantic role labeling. In *Proceedings of the IJCNLP, ACL '09*, 28–36. Stroudsburg, PA, USA: ACL.
- Datla, V. V.; Lin, K.-I.; and Louwerse, M. M. 2014. Part of speech induction from distributional features: Balancing vocabulary and context. In *Proceedings of the FLAIRS*.
- Garg, N., and Henderson, J. 2012. Unsupervised semantic role induction with global role ordering. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2, ACL '12*, 145–149. Stroudsburg, PA, USA: ACL.
- Grenager, T., and Manning, C. D. 2006. Unsupervised discovery of a statistical verb lexicon. In *Proceedings of EMNLP, EMNLP '06*, 1–8. Stroudsburg, PA, USA: ACL.
- Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. H. 2009. The weka data mining software: An update. *SIGKDD Explor. Newsl.* 11(1):10–18.
- Kingsbury, P., and Palmer, M. 2002. From treebank to propbank. In *Proceedings of LREC*.
- Lang, J., and Lapata, M. 2010. Unsupervised induction of semantic roles. In *Proceedings of ACL-HLT*, 939–947. ACL.
- Lang, J., and Lapata, M. 2011. Unsupervised semantic role induction via split-merge clustering. In *Proceedings of the ACL-HLT, HLT '11*, 1117–1126. Stroudsburg, PA, USA: ACL.
- Modi, A.; Titov, I.; and Klementiev, A. 2012. Unsupervised induction of frame-semantic representations. In *Proceedings of NAACL-HLT, WILS '12*, 1–7. Stroudsburg, PA, USA: ACL.
- Pradhan, S.; Hacioglu, K.; Krugler, V.; Ward, W.; Martin, J. H.; and Jurafsky, D. 2005. Support vector learning for semantic argument classification. *Mach. Learn.* 60(1-3):11–39.
- Socher, R.; Bauer, J.; Manning, C. D.; and Andrew Y., N. 2013. Parsing with compositional vector grammars. In *Proceedings of ACL*, 455–465. Sofia, Bulgaria: ACL.
- Solan, Z.; Horn, D.; Ruppín, E.; and Edelman, S. 2005. Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences of the United States of America* 102(33):11629–11634.
- Titov, I., and Klementiev, A. 2011. A bayesian model for unsupervised semantic parsing. In *Proceedings of ACL-HLT, HLT '11*, 1445–1455. Stroudsburg, PA, USA: ACL.